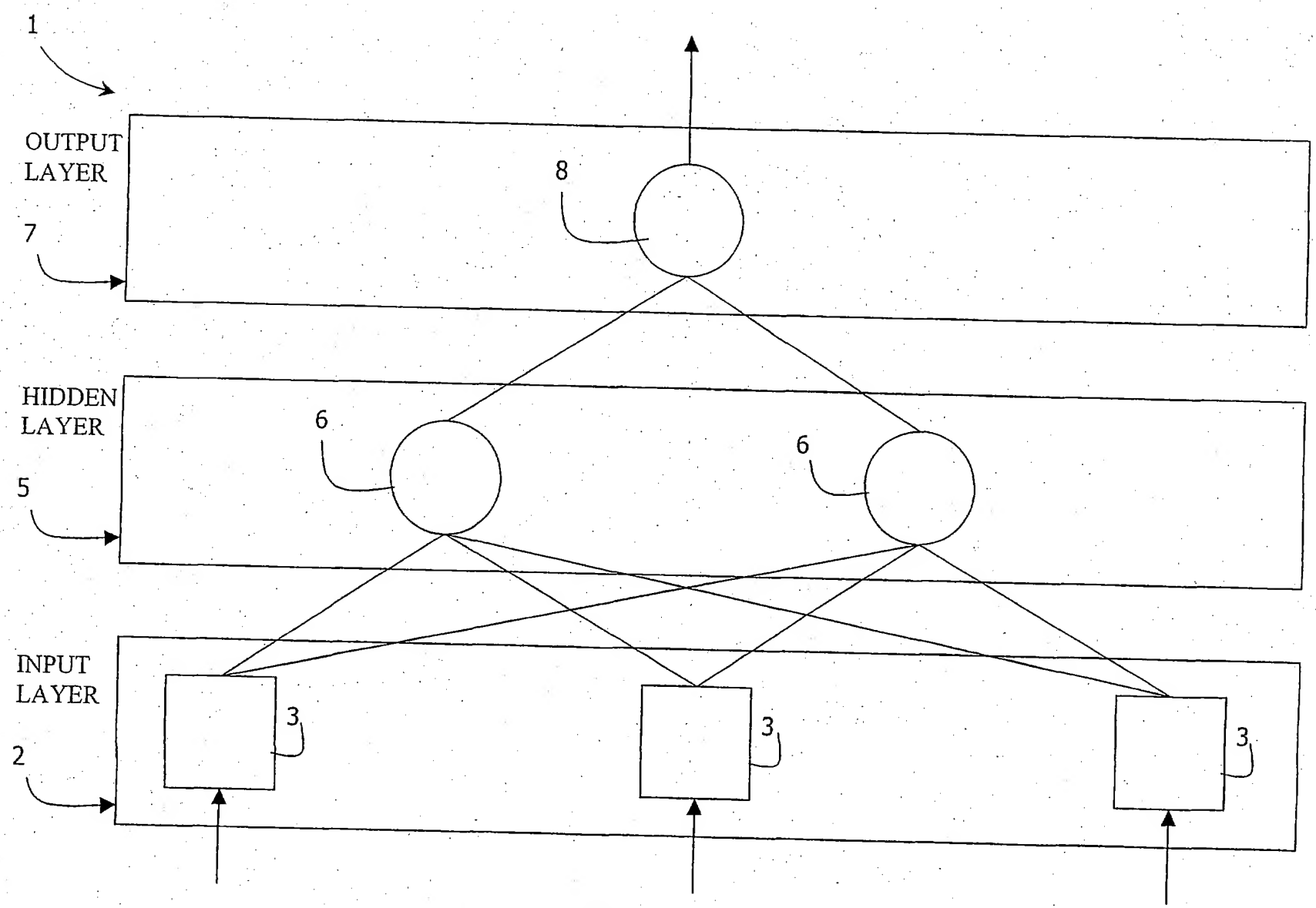


FIG. 1.

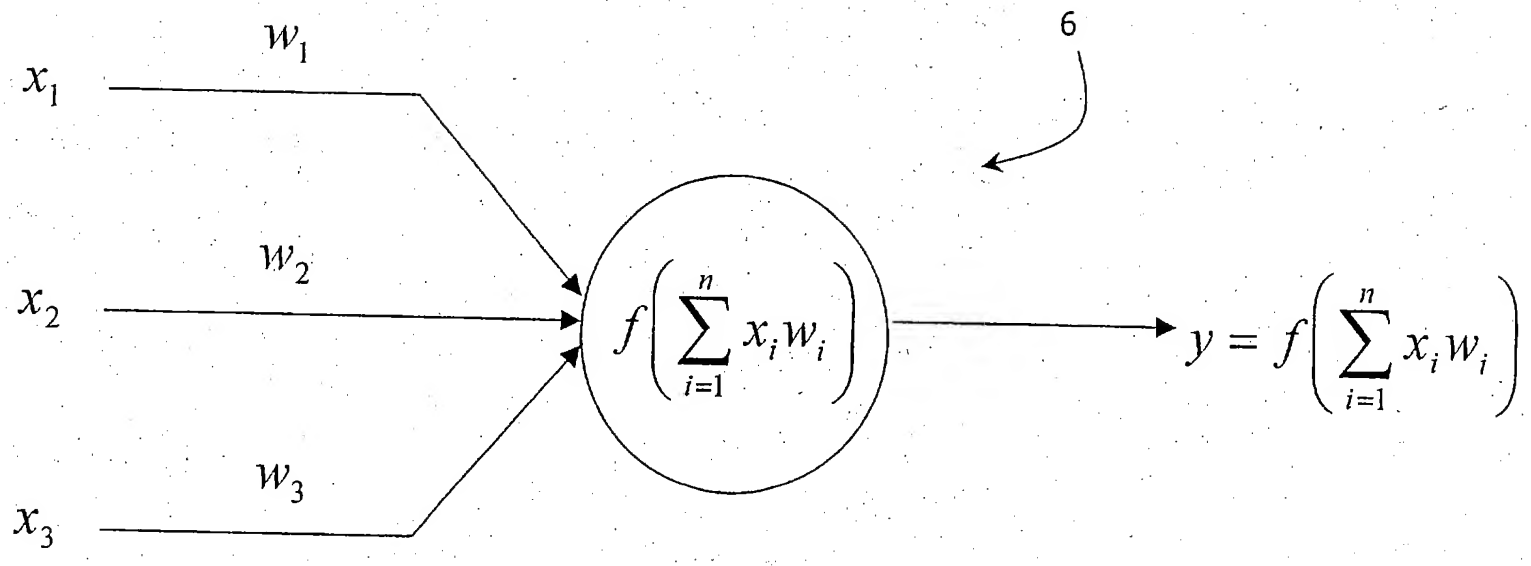
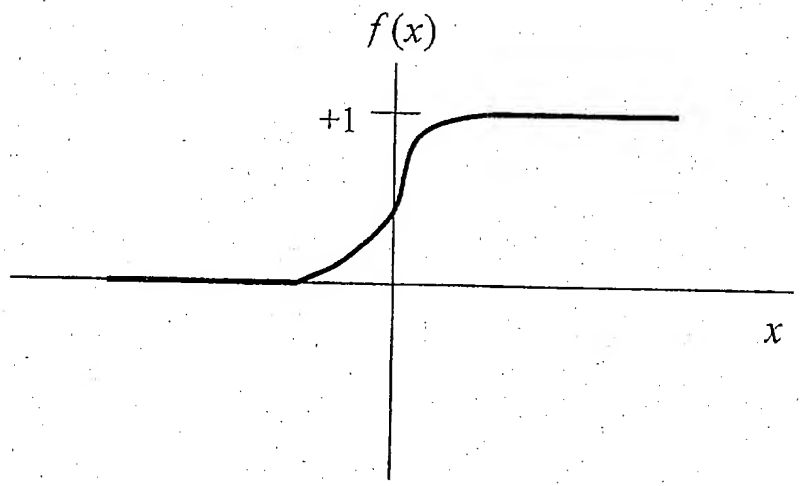

w_1
x_1
6
w_2
x_2
$f\left(\sum_{i=1}^{n} x_i w_i\right)$
$y = f\left(\sum_{i=1}^{n} x_i w_i\right)$
w_3
x_3

FIG. 2.



$$f(x) = \frac{1}{\left(1 + e^{-(x+T)}\right)}$$

FIG. 3.

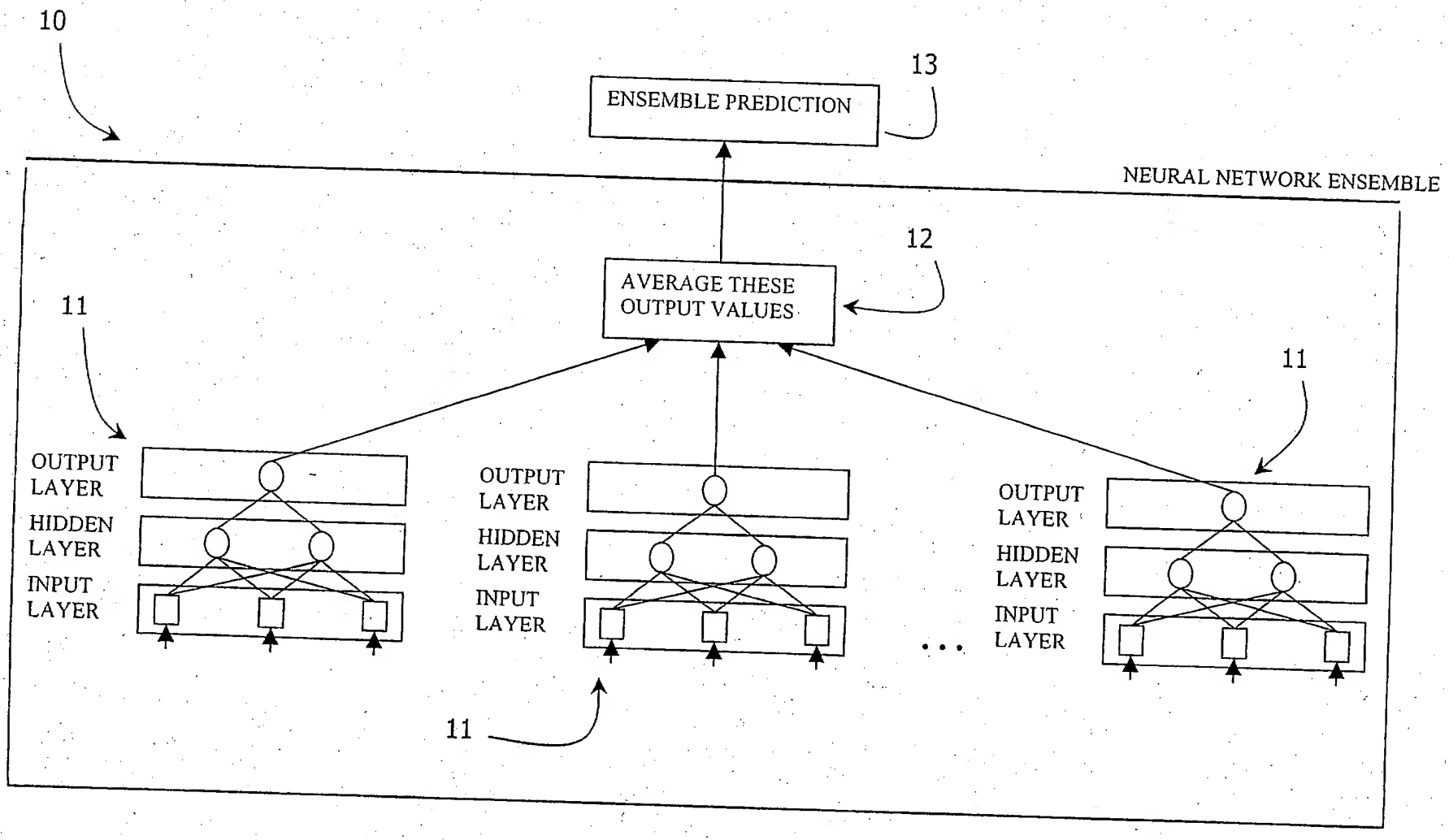

10
13
ENSEMBLE PREDICTION
NEURAL NETWORK ENSEMBLE
12
AVERAGE THESE OUTPUT VALUES
11
11
11
OUTPUT LAYER
HIDDEN LAYER
INPUT LAYER
OUTPUT LAYER
HIDDEN LAYER
INPUT LAYER
OUTPUT LAYER
HIDDEN LAYER
INPUT LAYER
...

FIG. 4.

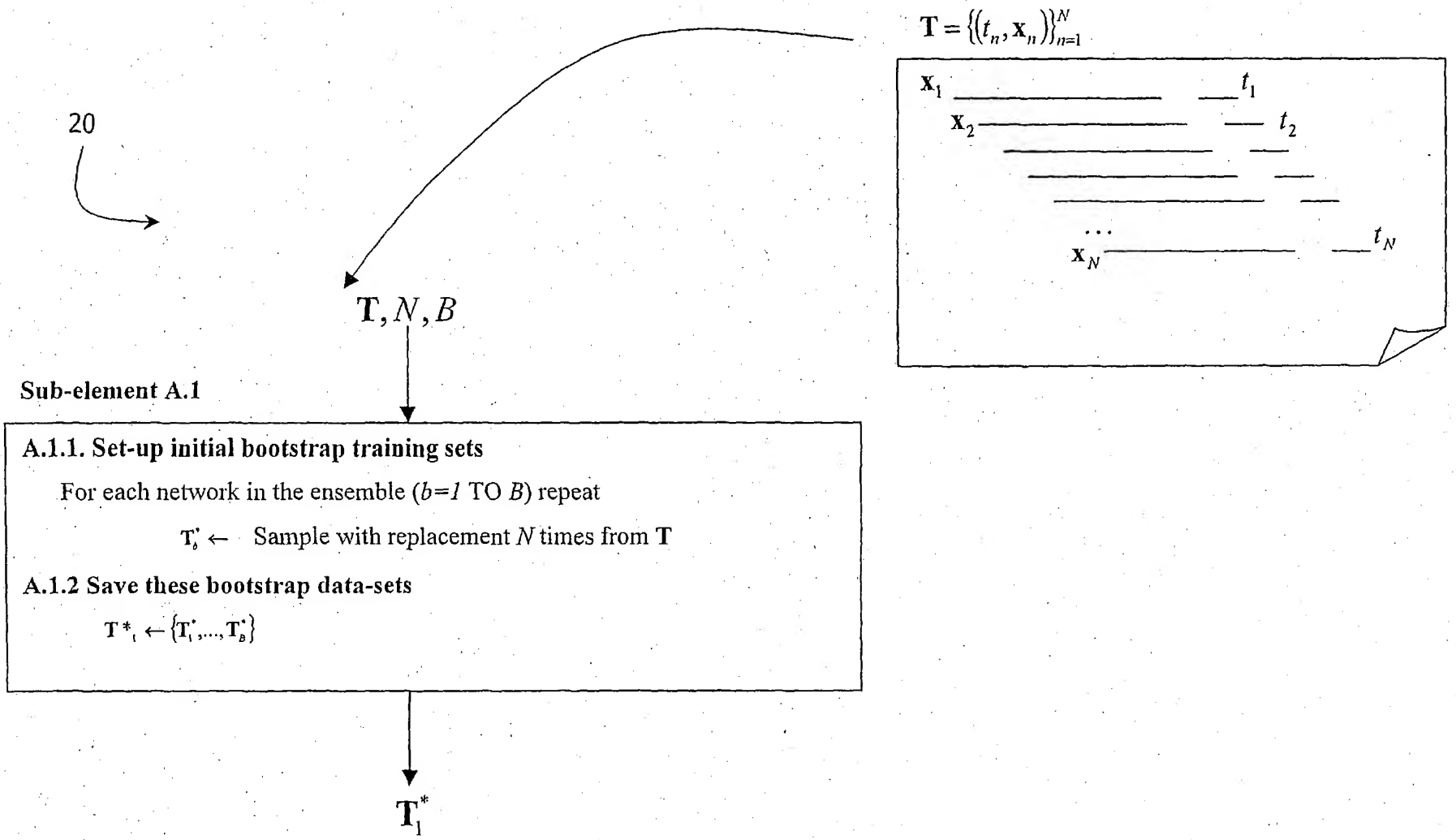

20
$\mathbf{T} = \{(t_n, \mathbf{x}_n)\}_{n=1}^{N}$
$\mathbf{x}_1$ t_1
$\mathbf{x}_2$ t_2
$\mathbf{x}_N$ t_N
$\mathbf{T}, N, B$
Sub-element A.1
A.1.1. Set-up initial bootstrap training sets
For each network in the ensemble (b=1 TO B) repeat
$\mathbf{T}_b^* \leftarrow$ Sample with replacement N times from $\mathbf{T}$
A.1.2 Save these bootstrap data-sets
$\mathbf{T}^*_t \leftarrow \{\mathbf{T}_1^*, \ldots, \mathbf{T}_B^*\}$
$\mathbf{T}_1^*$

FIG. 5.

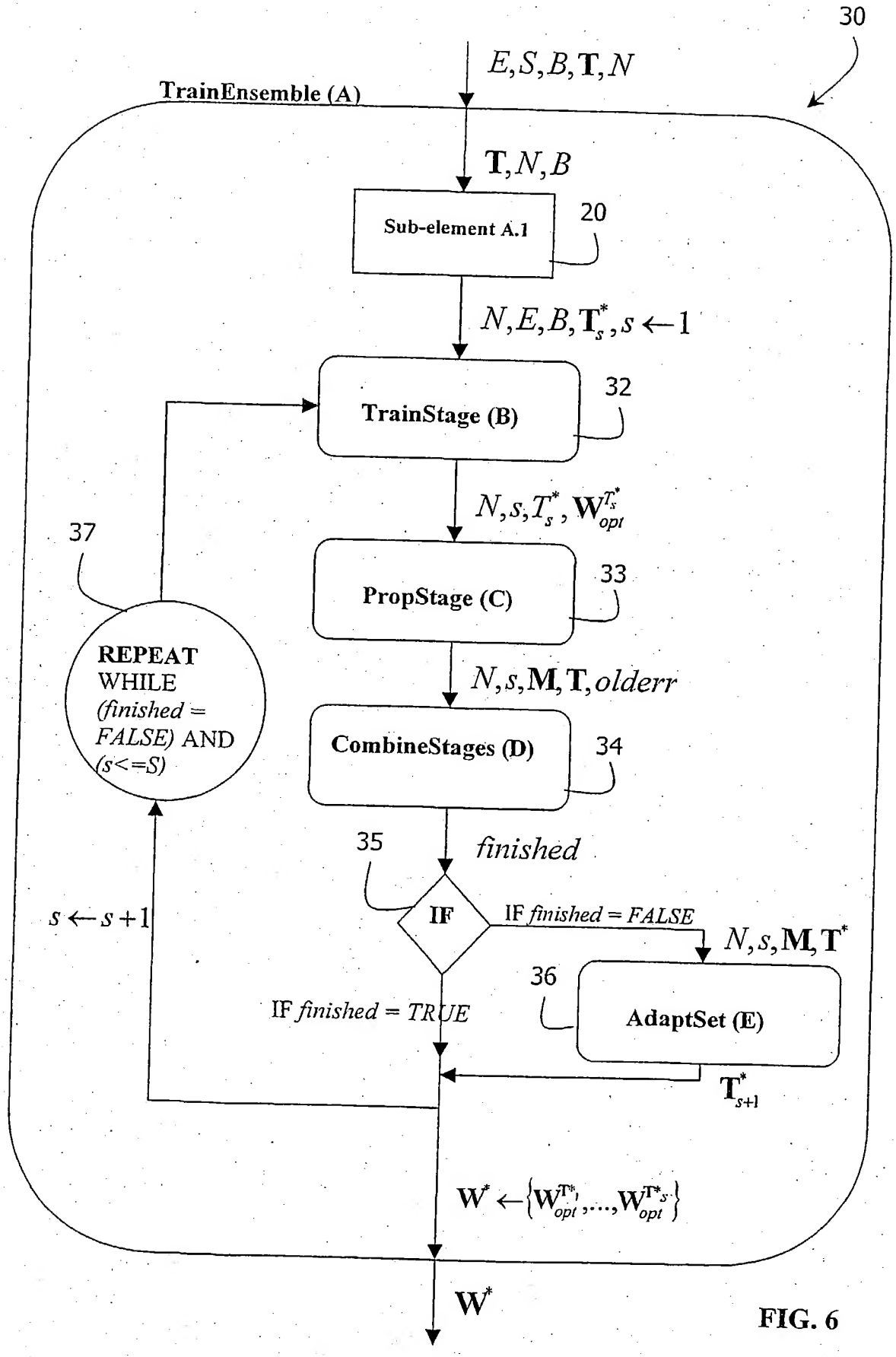

30
$E,S,B,\mathbf{T},N$
TrainEnsemble (A)
$\mathbf{T},N,B$
Sub-element A.1
20
$N,E,B,\mathbf{T}_s^*,s \leftarrow 1$
TrainStage (B)
32
$N,s,T_s^*,\mathbf{W}_{opt}^{T_s^*}$
37
PropStage (C)
33
REPEAT WHILE (finished = FALSE) AND (s<=S)
$N,s,\mathbf{M},\mathbf{T},olderr$
CombineStages (D)
34
35
finished
IF
$s \leftarrow s+1$
IF finished = FALSE
$N,s,\mathbf{M},\mathbf{T}^*$
36
IF finished = TRUE
AdaptSet (E)
$\mathbf{T}_{s+1}^*$
$\mathbf{W}^* \leftarrow \{\mathbf{W}_{opt}^{T_1^*},\ldots,\mathbf{W}_{opt}^{T_s^*}\}$
$\mathbf{W}^*$

FIG. 6

32

$N, E, S, B, \mathbf{T}_s^*$

TrainStage (B)

B.1 Copy training sets for this stage into individual sets

$\{\mathbf{T}_1^*, \ldots, \mathbf{T}_B^*\} \leftarrow \mathbf{T}_s^*$

B.2 Compute generalisation error estimates for each training vector

For every training vector (*n=1* TO *N)* in the original training set repeat

For every epoch (*e=1* TO *E*) repeat

Compute:

$$\mathbf{G}_e^n \leftarrow \left(t_n - \frac{\sum_{b=1}^{B} \gamma_n^b \left(\phi\left(\mathbf{x}_n ; \mathbf{w}_e^{\mathbf{T}_b^*} \right) \right)}{\sum_{b=1}^{B} \gamma_n^b} \right)^2$$

B.3 Aggregate the ensemble generalisation error estimates

For every epoch (*e=1* TO *E*) repeat

Compute:

$$\mathbf{A}_e \leftarrow \frac{1}{N} \sum_{n=1}^{N} \mathbf{G}_e^n$$

B.4 Find the best value for *e* for each network in the ensemble

$$e_{opt} \leftarrow \arg\min_e (\mathbf{A}_e)$$

$$\mathbf{W}_{opt}^{\mathbf{T}_s^*} \leftarrow \mathbf{w}_{e_{opt}}^{\mathbf{T}_s^*}$$

$\mathbf{W}_{opt}^{\mathbf{T}_s^*}$

FIG. 7

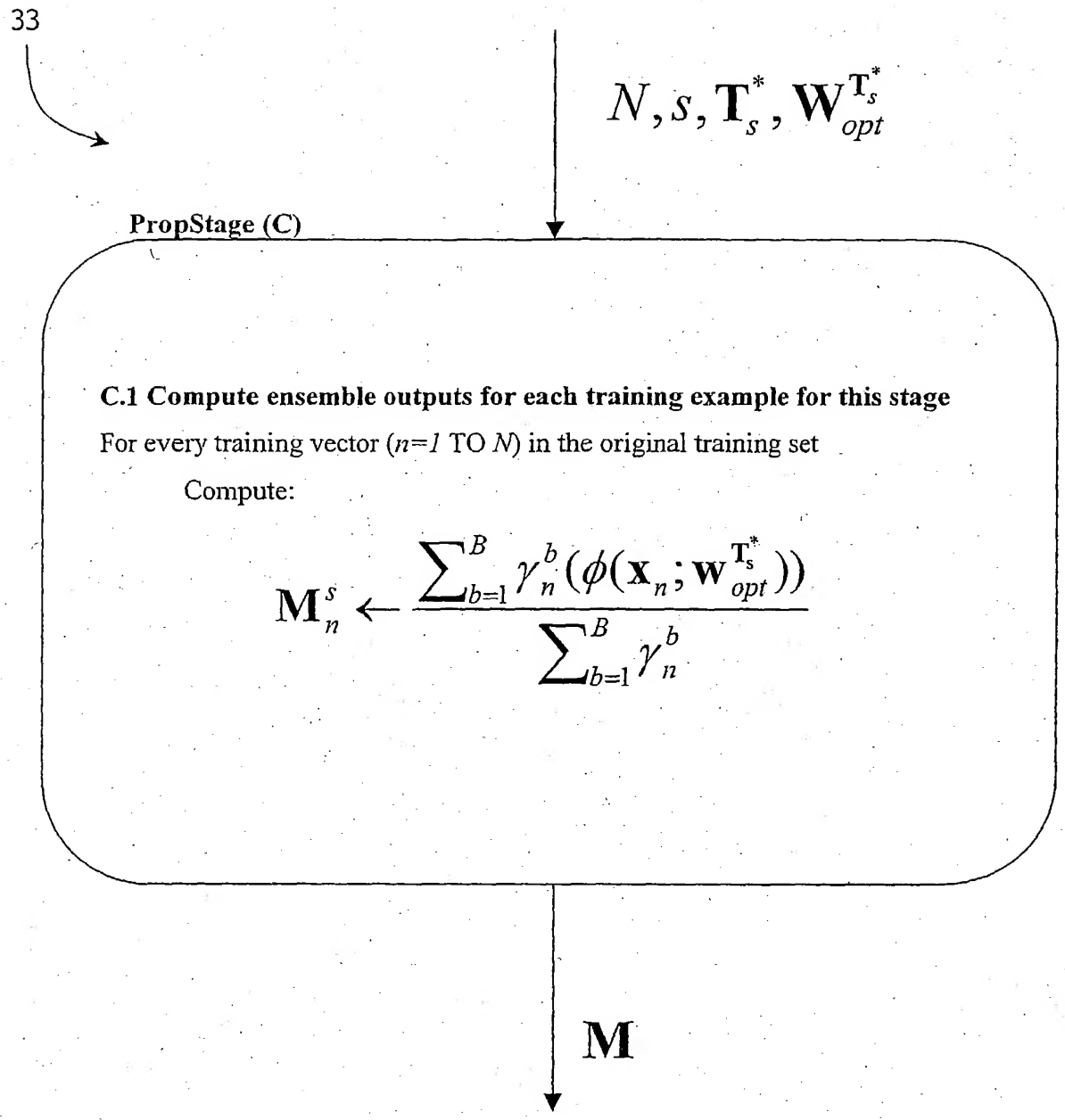

33
$N, s, \mathbf{T}_s^*, \mathbf{W}_{opt}^{\mathbf{T}_s^*}$
PropStage (C)
C.1 Compute ensemble outputs for each training example for this stage
For every training vector (n=1 TO N) in the original training set
Compute:
$\mathbf{M}_n^s \leftarrow \frac{\sum_{b=1}^{B} \gamma_n^b (\phi(\mathbf{x}_n; \mathbf{w}_{opt}^{\mathbf{T}_s^*}))}{\sum_{b=1}^{B} \gamma_n^b}$
$\mathbf{M}$

FIG. 8

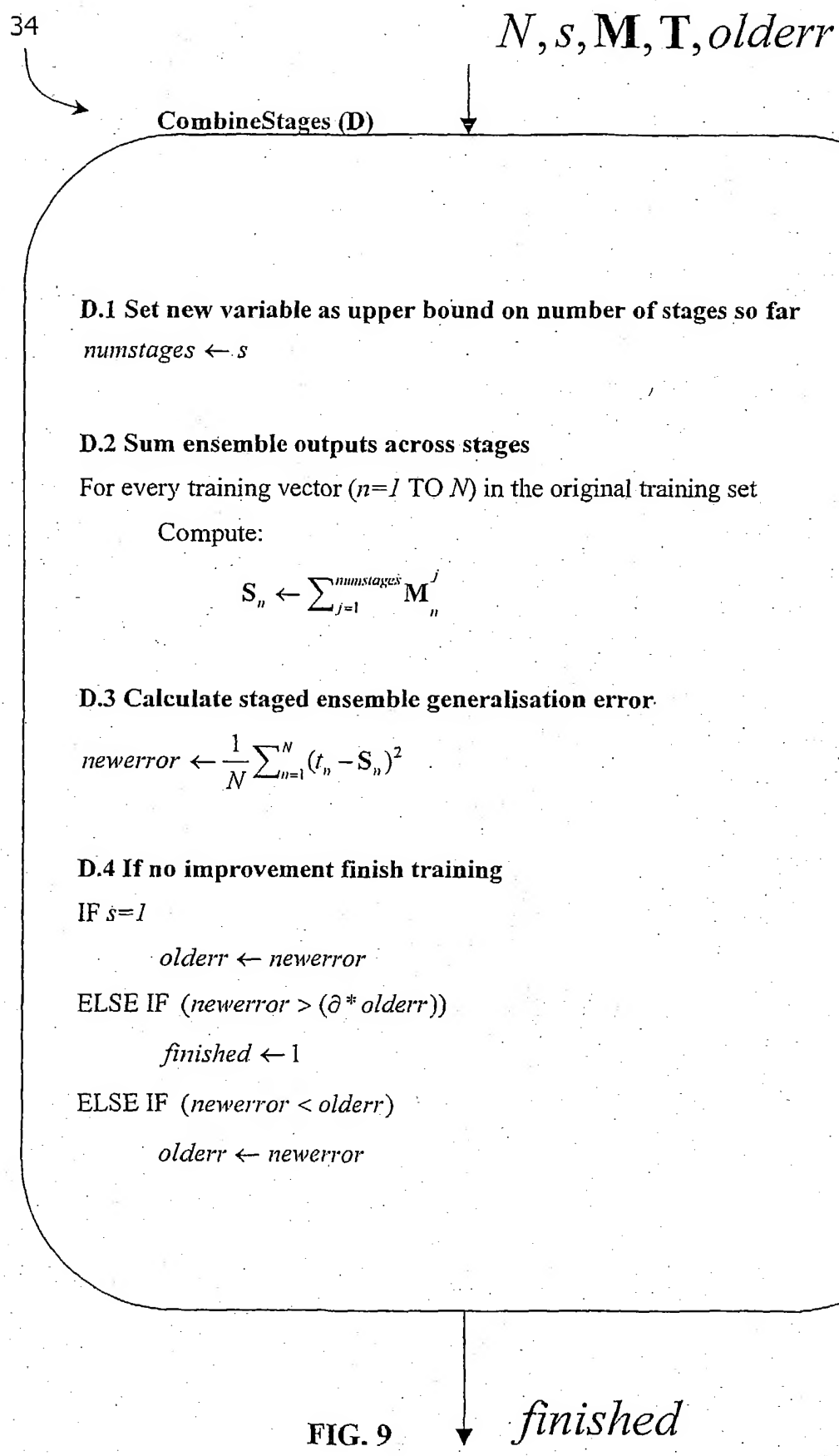

34
$N, s, \mathbf{M}, \mathbf{T}, olderr$
CombineStages (D)
D.1 Set new variable as upper bound on number of stages so far
$numstages \leftarrow s$
D.2 Sum ensemble outputs across stages
For every training vector ($n=1$ TO N) in the original training set
Compute:
$\mathbf{S}_n \leftarrow \sum_{j=1}^{numstages} \mathbf{M}_n^j$
D.3 Calculate staged ensemble generalisation error
$newerror \leftarrow \frac{1}{N} \sum_{n=1}^{N} (t_n - \mathbf{S}_n)^2$
D.4 If no improvement finish training
IF $s=1$
$olderr \leftarrow newerror$
ELSE IF ($newerror > (\partial * olderr)$)
$finished \leftarrow 1$
ELSE IF ($newerror < olderr$)
$olderr \leftarrow newerror$
$finished$

FIG. 9

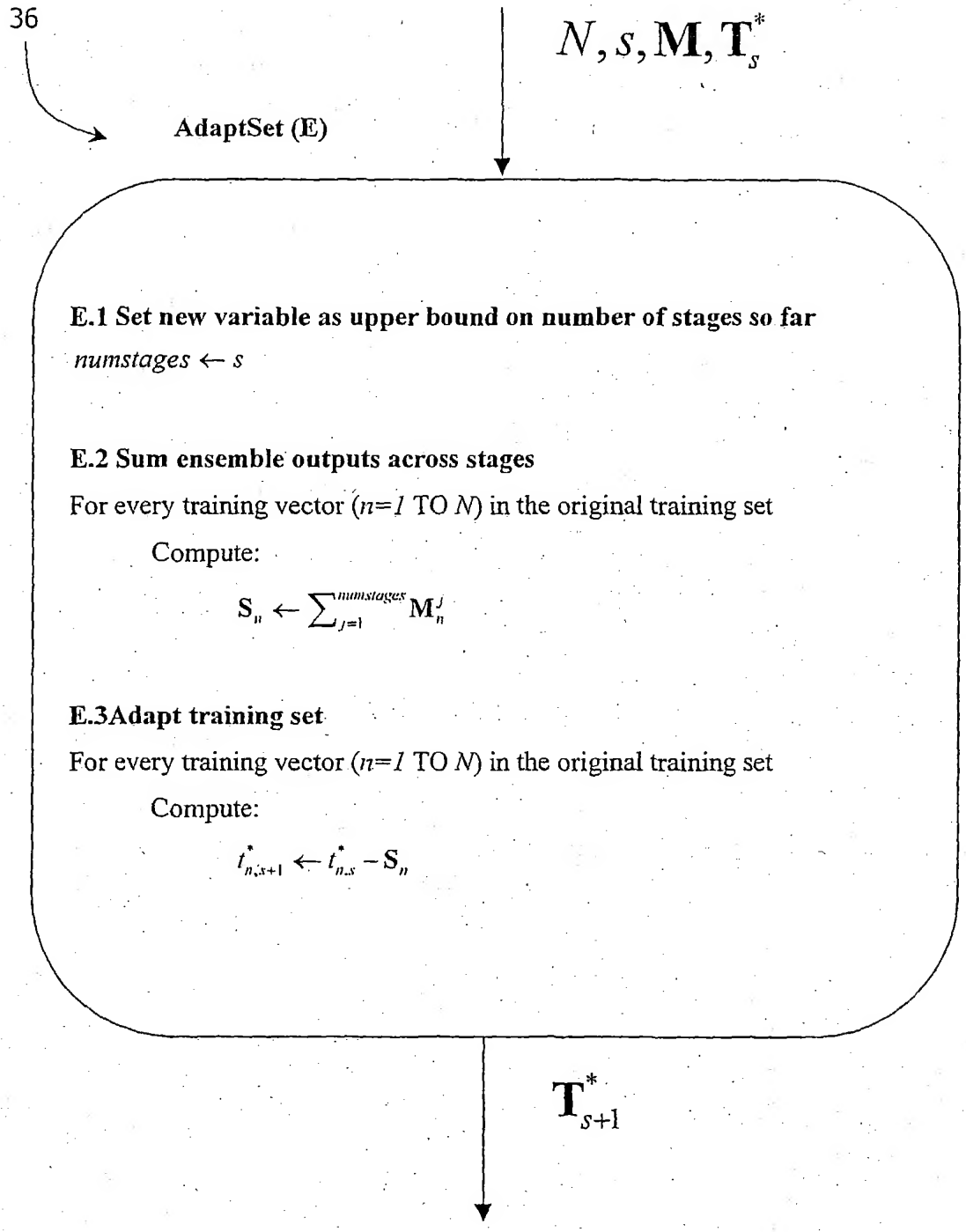

36
$N, s, \mathbf{M}, \mathbf{T}_s^*$
AdaptSet (E)
E.1 Set new variable as upper bound on number of stages so far
numstages ← s
E.2 Sum ensemble outputs across stages
For every training vector (n=1 TO N) in the original training set
Compute:
$\mathbf{S}_n \leftarrow \sum_{j=1}^{numstages} \mathbf{M}_n^j$
E.3Adapt training set
For every training vector (n=1 TO N) in the original training set
Compute:
$t_{n,s+1}^* \leftarrow t_{n,s}^* - \mathbf{S}_n$
$\mathbf{T}_{s+1}^*$

FIG. 10